Published in final edited form as:

Cancer Res. 2016 December 1; 76(23): 7012–7023. doi:10.1158/0008-5472.CAN-16-1371.

Genetic polymorphisms in the long noncoding RNA MIR2052HG offer a pharmacogenomic basis for the response of breast cancer patients to aromatase inhibitor therapy

James N. Ingle[1,*], **Fang Xie**[2,*], **Matthew J. Ellis**[3,*], **Paul E. Goss**[4], **Lois E. Shepherd**[5], **Judith-Anne W. Chapman**[5], **Bingshu E. Chen**[5], **Michiaki Kubo**[6], **Yoichi Furukawa**[7], **Yukihide Momozawa**[6], **Vered Stearns**[8], **Kathleen I. Pritchard**[9], **Poulami Barman**[10], **Erin E. Carlson**[10], **Matthew P. Goetz**[1], **Richard M. Weinshilboum**[2], **Krishna R. Kalari**[10], and **Liewei Wang**[2]

[1]Division of Medical Oncology, Mayo Clinic, Rochester, MN

[2]Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN

[3]Baylor Breast Center, Houston, TX

[4]Massachusetts General Hospital Cancer Center, Harvard University, Boston, MA

[5]Canadian Cancer Trials Group, Kingston, Ontario, Canada

[6]RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan

[7]University of Tokyo, Tokyo, Japan

[8]Johns Hopkins School of Medicine, Baltimore, MD

[9]Sunnybrook Odette Regional Cancer Centre, University of Toronto, Toronto, Ontario, Canada

[10]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN

Abstract

Genetic risks in breast cancer remain only partly understood. Here we report the results of a genome-wide association study of germline DNA from 4,658 women, including 252 women experiencing a breast cancer recurrence, who were entered on the MA.27 adjuvant trial comparing the aromatase inhibitors (AI) anastrozole and exemestane. Single nucleotide polymorphisms (SNP) of top significance were identified in the gene encoding MIR2052HG, a long noncoding RNA of unknown function. Heterozygous or homozygous individuals for variant alleles exhibited a ~40% or ~63% decrease, respectively, in the hazard of breast cancer recurrence relative to homozygous wild-type individuals. Functional genomic studies in lymphoblastoid cell lines and ERα-positive breast cancer cell lines showed that expression from MIR2052HG and the ESR1 gene encoding estrogen receptor-α (ERα) was induced by estrogen and AI in a SNP-dependent

Correspondence and reprint requests: James N. Ingle, M.D., Mayo Clinic, 200 First Street S.W., Rochester, MN 55905, ingle.james@mayo.edu, phone: 507-284-4857, fax 507-284-1803.

*Contributed equally

Conflict of interest: No potential conflicts of interest were disclosed by the authors.

manner. Variant SNP genotypes exhibited increased ERα binding to estrogen response elements, relative to wild-type genotypes, a pattern that was reversed by AI treatment. Further, variant SNPs were associated with lower expression of MIR2052HG and ERα. RNAi-mediated silencing of MIR2052HG in breast cancer cell lines decreased ERα expression, cell proliferation and anchorage-independent colony formation. Mechanistic investigations revealed that MIR2052HG sustained ERα levels both by promoting AKT/FOXO3-mediated ESR1 transcription and by limiting ubiquitin-mediated, proteasome-dependent degradation of ERα. Taken together, our results define MIR2052HS as a functionally polymorphic gene that affects risks of breast cancer recurrence in women treated with AI. More broadly, our results offer a pharmacogenomic basis to understand differences in the response of breast cancer patients to AI therapy.

Keywords

GWAS; breast cancer; aromatase inhibitors; functional genomics; *MIR2052HG*

Introduction

Breast cancer is the most common form of cancer in women both in the United States (1) and worldwide (2). Endocrine therapy is the most important treatment modality in the majority of women who have estrogen receptor α (ERα) positive breast cancer. Whereas tamoxifen, a selective ER modulator (SERM), has substantial value in reducing the risk of disease recurrence in women with ERα-positive early stage breast cancer (3), a recent meta-analysis demonstrated that, when compared directly as monotherapy, aromatase inhibitors (AIs) were superior to tamoxifen in terms of local recurrence, distant recurrence, contralateral recurrence, breast cancer mortality, and all-cause mortality (4). However, despite the clear efficacy of AIs as adjuvant therapy, this meta-analysis revealed that 19.1% of women treated with an AI, anastrozole or letrozole, experienced a recurrence of their breast cancer at 10 years and there was no indication of a plateau in the time to recurrence curve (4).

Canadian Cancer Trials Group MA.27 (5) is the largest adjuvant endocrine therapy trial that exclusively studied AIs. Postmenopausal women with hormone receptor-positive early stage breast cancer were randomized to the steroidal AI exemestane or the non-steroidal AI anastrozole, and no difference in efficacy was identified (5). We performed a genome-wide association study (GWAS), which indicated that germline genetic variability in SNP genotypes related to a gene encoding a lncRNA, may alter ERα expression and impact outcomes after treatment with AIs.. Functional genomic studies of this lncRNA and the SNPs related to it provided novel mechanisms by which the lncRNA might affect the level of AI benefit.

Methods

Source of Patients

Patients were obtained from the MA.27 trial (ClinicalTrials.gov number NCT00066573). MA.27 included postmenopausal women with histologically confirmed and completely

resected stage I–III breast cancer (AJCC Version 6) that was ERα and/or PgR positive. Patients were randomized to five years of anastrozole or exemestane. Only North American patients were offered participation in collection of blood specimens and 5221 of 6827 (76.5%) of the North American patients contributed blood and gave consent for genetic testing. This research was performed after approval by local institutional review boards in accordance with assurances filed with, and approved by, the Department of Health and Human Services.

Primary Outcome: Breast Cancer-Free Interval

The primary outcome was the STEEP endpoint of BCFI (6), defined as the time from randomization to the first local-regional breast cancer recurrence (including ipsilateral DCIS), distant breast cancer recurrence, contralateral breast cancer (invasive or DCIS) or death with or from breast cancer without a prior recurrence date. Follow-up was censored at non-breast cancer death, or longest follow-up without recurrence.

Genotyping, Imputation, and Quality Control

Three cohorts of patients from MA.27 were genotyped by the RIKEN Center for Integrative Medical Sciences. Cohort 1 involved patients genotyped as part of a GWAS with musculoskeletal adverse events as the phenotype utilizing the Illumina Human610 Quad Beadchip (7). Cohort 2 involved patients genotyped as part of a GWAS with fragility bone fractures as the phenotype utilizing the Illumina Human OmniExpress platform (8). Cohort 3 involved the remainder of the patients from MA.27 with DNA and consent. The quality control measures for cohorts 1 and 2 have been published (7, 8). For cohort 3, the following measures were taken for quality control purposes. One case and three controls were randomly chosen as duplicates for quality control of genotype concordance. A Caucasian parent-child Centre d'Etude du Polymorphisme Human trio from the HapMap was included to check for Mendelian transmission of alleles. Genotypes were determined utilizing the Illumina Human OmniExpressExome platform. This platform provided genotype data for 964,193 SNPs of which 2,923 were removed because they were from chromosome Y, mitochondria, or unplaced chromosomes. Additionally, 40,631 SNPs failed genotyping and 250,843 were rare SNPs with MAF<0.01. Imputation was performed using EZimputer (9) across the three cohorts separately. We then combined the genotyping data from all three cohorts to perform the current GWAS. Since each of the three cohorts were genotyped with a different platform, there were some SNPs that were genotyped in some patients but not in others. EZimputer (9) (http://www.mayo.edu/research/departments-divisions/department-health-sciences-research/division-biomedical-statistics-informatics/software/bioinformatics-software-packages) imputes both un-genotyped SNPs and missing SNPs on a given platform. SNPs selected from imputation by EZimputer (9) had a dosage r squared >0.8 [r squared here is defined as the estimated squared correlation between the estimated allele dosage (0*P(Hom Ref/first) + 1*P(AB) + 2*P(Hom Alt/second)) and the true allele dosage].

The data from this GWAS have been deposited in the Data Base of Genotypes and Phenotypes (dbGaP). The dbGaP Study Accession Number is phs001043 and the URL is http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001043.v1.p1.

Deep Sequencing Methodology

We performed deep sequencing of the region containing the top imputed SNPs to determine the quality of the imputation as detailed in Supplementary Materials.

Statistical Analyses

Exact Fisher tests were used to examine whether there were imbalances between North American patients who were included in this GWAS and those who were not. We also evaluated if any additional clinical variables were significantly associated with breast recurrence event using Stepwise selection method (10). Our analysis employed a stratified genome-wide Cox-proportional hazards model using significant stratification factors and the model was further controlled for additional covariates including treatment arm, cohort, race, ER/PgR status, T-stage, ECOG performance score and bisphosphonate use. To avoid biases that might arise from differences in genetic ancestry (i.e., population stratification), the EigenStrat software was used to determine eigenvalues for the SNP correlation matrix that statistically differed from zero based on Tracy-Widom p-values (11) (12). All the analyses were run using the R statistical computing package (13), PLINK (14), and SAS (SAS Institute, Cary, NC).

Haplotype analysis was performed using the top six SNPs. Haplotype probabilities for individual samples were estimated using haplo.stats package v1.7.7 in R v3.2.0 (15) (details in Supplementary Materials)

Deep Sequencing Data Analysis

In order to determine the quality of imputation, we performed targeted deep sequencing of the 300kb region (chr8: 75,400,000–75,700,000) surrounding the chromosome 8 GWAS signal in a total of 997 patients (249 with and 748 without a breast event). We removed both 5′ and 3′-end primers with cutadapt-1.7.1 with minimum-length option = 20. The trimmed reads were aligned to the hg19 reference genome using BWA-MEM. SNVs and INDELs were called using HaplotypeCaller and genotyped across all samples with GenotypeGVCF from the Genome Analysis Tool Kit (GATK) (16–18). Variants were annotated with functional features, impact prediction, and clinical significance using SnpEff (19), ClinVar, HGMD, and ExAC population frequencies with the BioR annotation tool (20). To demonstrate the imputation quality of our GWAS SNPs, we correlated the variant calls from deep sequencing with a set of imputed SNPs from the GWAS signal.

Cell lines

Human breast cancer cell lines CAMA-1, HCC1428, BT474, AU565, BT549 and human embryonic kidney cell line 293T were obtained from American Type Culture Collection (ATCC, Manassus, VA) in 2014 and the identities of all cell lines were confirmed by the medical genome facility at Mayo Clinic Center (Rochester MN) using short tandem repeat profiling upon receipt. The breast cancer cell line MCF7/AC1, stably overexpressing aromatase (stably transfected *CYP19A1* gene), was generously gifted from Angela H. Brodie, Ph. D. (University of Maryland, Baltimore, MD). The cells were authenticated in 2015 by Genetica DNA Laboratories (Cincinnati, OH) using a StemElite ID system that uses

short tandem repeat genotyping. All the cells used in our studies were within the initial five passages.

Functional Genomics Studies

Details with respect to materials used, cell culture and lymphoblastoid cell line (LCLs) techniques, quantitative real-time PCR assay (qRT-PCR), chromatin Immunoprecipitation (ChIP) assays, cell proliferation assays, colony forming assays, western blotting are given in the Supplementary Materials.

Results

Population of Patients Studied

The Participant Flow Diagram (Supplementary Fig. S1) shows the patients included and excluded from the GWAS, and Supplementary Table 1 provides the patient characteristics and analyses revealed good comparability (Supplementary Materials).

A total of 4,784 patients had sufficient DNA for genotyping but 4,658 patients were utilized for the GWAS following quality control procedures as specified in the Supplementary Materials and in the Participant Flow Diagram (21) (Supplementary Fig. S1). The analysis included 252 women with a breast event and 4406 women who had not experienced a breast event (Table 1). We did not find any eigenvectors that were significantly associated with BCFI.

Table 1 shows that the percentage of patients with a breast event was similar within each of the three genotyped cohorts (Cohort 1: 4.9%, Cohort 2: 5.3%, Cohort 3: 5.6%). Patients with and without a breast event were well balanced for age, treatment arm (anastrozole versus exemestane), and baseline body mass index. There were significant imbalances for adjuvant chemotherapy (P=3.68e-12), T stage (P=<2e-16), lymph node status (p=<2e-16) Eastern Cooperative Oncology Group (ECOG) performance score (P=0.0003), bisphosphonate use (P=0.0004), and ER/progesterone receptor (PgR) (p=0.008) with patients with a breast event having a higher percentage of ER-positive/PgR-negative and a lower percentage of ER-positive/PgR-positive tumors.

Genotyping and Imputation

Genotyping for Cohorts 1 and 2 has been described previously (7, 8) utilizing the Illumina Human610 Quad BeadChip and the Illumina HumanOmniExpress platforms, respectively. For Cohort 3, after quality control measures, 669,796 genotyped SNPs were available for combining with genotype data from Cohorts 1 and 2. Imputation was performed using an in-house method, EZimputer (9), which returned a total of 9.57M SNPs (observed plus imputed) with r squared >0.8, of which 7.4M (observed + imputed) SNPs had a MAF>0.01 and were used for the analyses.

Genome-wide Association Study Analyses

We performed a stratified Cox-proportional hazards analysis utilizing stratification factors and other covariates as detailed in the Methods section. The Manhattan plot (Fig. 1A) shows

that the SNPs with the lowest p-values mapped to chromosome 8 and the quantile-quantile plot (Supplementary Fig. S2) revealed a lambda of 0.999. Figure 1B shows the locus zoom for the region with the top SNPs. Characteristics of the top six SNPs on chromosome 8 are listed in Table 2 and these SNPs are all "favorable" in that they were associated with longer BCFI. All six SNPs were in strong linkage disequilibrium with R^2 values ranging from 0.95 to 0.99. The hazard ratios provided in Table 2 are for the presence of a single variant SNP genotype, which would be a heterozygous state. The presence of two variant SNP genotypes would be a homozygous variant state and the hazard ratio would be multiplicative. For example, considering rs4476990, the hazard ratio was 0.61, indicating ~39% reduction in the hazard of a breast cancer event for the heterozygous state relative to the homozygous wild type state, but the hazard ratio for women with the homozygous variant state would be 0.61×0.61=0.37, indicating ~63% reduction in the hazard of a breast cancer event, again relative to those with a homozygous wild type state. The p-values for the top six SNPs ranged from 2.15E-07 to 6.24E-07). Importantly, these variant SNP genotypes were common with MAFs ranging from 0.32 to 0.42.

To validate our imputation quality, we used deep sequencing techniques to call variants surrounding the chromosome 8 peak. We compared the deep sequencing variants calls of these SNPs with those obtained from MA-27 data. The correlations between the variants (from deep sequencing) and the MA.27 SNPs were at least 0.9 in all the top SNPs (Supplementary Table 2), and thus of high quality.

The top SNP (rs13260300) was located 32.4kb 5′ of *MIR2052HG* (other names: *FLJ39080*, LOC441355), a gene located on chromosome 8q21.11, and two of the top SNPs were located in *MIR2052HG. MIR2052HG* encodes a lncRNA whose function is not known and with very few publications referring to this gene (22, 23).

Because this study involved AIs as therapy for ER-positive breast cancer, we interrogated the top SNPs to determine which were located in or within 500 bp of a putative estrogen response element (ERE), similar to what we have done in previous studies (24, 25). The distance of 500 bp was chosen as chromatin immunoprecipitation (ChIP) assays become less reliable with greater distances between a SNP and an ERE. Two of the top SNPs fulfilled one of these two criteria with rs4476990 being located 33kb 5′ of *MIR2052HG* and in an ERE, while rs3802201 mapped to intron 1 of *MIR2052HG* and located 16bp from an ERE. We focused on these two SNPs, which are in moderate linkage disequilibrium (R^2=0.6), in our functional studies. The significant relationships between the genotypes of the two SNPs, rs4476990 and rs3802201, and BCFI are shown in Supplementary Figs. 3A and 3B, respectively. After adjustment for covariates (treatment arm, cohort, race, ER/PR status, T-Stage, ECOG performance status and bisphosphonate use), a stratified Cox model using stratification factors (adjuvant chemotherapy, lymph node status, trastuzumab use) determined p=2.51E-07 for rs4476990 and p=6.24E-07 for rs3802201.

Haplotype Analysis of the Top Six SNPs

A haplotype analysis of the top six SNPs was performed to determine the degree of association. The results of this analysis are shown in Supplementary Table 3 and reveal that

the association for the most significant haplotype (HR 0.58, p=1.23E-06) was not superior to the top SNPs considered individually (Table 2).

Functional Genomic Studies of Chromosome 8 SNPs

Because of the importance of ERα in endocrine therapy, we interrogated the Cancer Genome Atlas (TCGA) breast cancer data (26) for a possible relationship between the expression of MIR2052HG and ESR1, which encodes ERα, in 485 ER-positive breast cancers. There was a positive correlation (Spearman correlation coefficient 0.370) between the expressions of these two genes. This provided an indication that there might be an important relationship between MIR2052HG and ESR1, which was supported by the functional genomic studies described in subsequent paragraphs. MIR2052HG is expressed in multiple breast cancer cell lines including ERα-positive cell lines, according to the Cancer Genomics Hub (https://cghub.ucsc.edu).

MIR2052HG SNPs determine estradiol-dependent MIR2052HG expression and ERα binding to EREs

As mentioned earlier, the rs4476990 SNP was located in a putative ERE and the rs3802201 SNP in intron 1 of *MIR2052HG* was near another ERE (Fig. 2A). To test the possible functional impact of these two SNPs, we utilized a model system consisting of 300 individual human lymphoblastoid cell lines (LCLs) for which we have extensive genomic and transcriptomic data. This LCL model system has repeatedly shown its value in both generating and testing pharmacogenomics hypotheses (8, 25, 27). Specifically, we selected 5 LCLs with homozygous wild type (WT) genotypes for both SNPs and 5 LCLs homozygous for variant genotypes for both SNPs to perform estradiol (E2) treatment and ChIP assays. In the presence of E2, cells homozygous for the variant SNP genotypes showed a dose-dependent increase in MIR2052HG expression (Fig. 2B) as well as increased binding of ERα to the EREs shown in Fig. 2A for variant genotypes for both SNPs relative to WT in ChIP assays using ERα antibody (Figs. 2C and 2D). 4-Hydroxy-tamoxifen (4-OH-Tam), a SERM that competes with E2 for ERα binding, could reverse this effect (Figs. 2C and 2D).

Aromatase inhibitors reverse estradiol-dependent and SNP-dependent MIR2052HG and ESR1 expression

The major function of AIs is to reduce estrogen levels by the inhibition of aromatase, the rate-limiting step in estrogen biosynthesis. The reduction of estrogens could have an effect on ERα-mediated function. As we have shown in our previous studies, estrogens and SERMs can alter gene expression in a SNP-dependent fashion (25). Thus, we proceeded to examine the effect of the AIs anastrozole and exemestane on MIR2052HG and ESR1 expression in the presence of androstenedione, which is aromatized to estrone by aromatase, the target for the AIs under study.

In the presence of androstenedione, LCLs with variant genotypes for both SNPs showed dose-dependent increases in MIR2052HG expression (Figs. 3A and B) that was similar to that for E2 (Fig. 2B). However, the addition an AI, either exemestane (Fig. 3A) or anastrozole (Fig. 3B), to the androstenedione-treated LCLs caused a “reversal” of the expression pattern with increased MIR2052HG expression in LCLs homozygous for the WT

genotypes but a marked decrease in LCLs homozygous for variant genotypes. Of particular interest was the observation of a direct correlation between this striking pattern of expression for MIR2052HG and that of ESR1 in the same cell lines (Figs. 3C and 3D), bringing us back to the correlation that we had observed between the expression of MIR2052HG and ESR1 in the TCGA data.

MIR2052HG affects ESR1 and ERα expression and proliferation, colony formation, and response to AIs in ERα-positive breast cancer cell lines

Having determined that, in a SNP and AI-dependent fashion, the expression of MIR2052HG was correlated with that of ESR1 (Fig. 3), we set out to study the possible functional impact of the MIR2052HG lncRNA on AI response and on cell proliferation. When we began our studies, the function of MIR2052HG was not known, but we hypothesized, based on the results shown in Fig. 3, that this lncRNA might influence AI response through its effect on the downstream expression of ESR1. To determine the effect of MIR2052HG knock down on ERα levels, we chose two ERα-positive breast cancer cell lines with relatively high endogenous MIR2052HG expression, CAMA1 and the aromatase expressing MCF7/AC1 (28) cell lines. Knock down of MIR2052HG resulted in striking decreases of ERα expression, both at the mRNA and the protein levels (Figs. 4A and 4B), consistent with the TCGA data that showed a positive correlation between the two genes. Furthermore, in MCF7/AC1 cells, knock down of MIR2052HG decreased cell proliferation and colony formation (Figs. 4C and 4D), while overexpression of MIR2052HG increased cell proliferation, colony formation, and ERα expression (Fig. 4E), functionally confirming the positive relationship between ERα and MIR2052HG. We observed the same results in two additional ERα-positive breast cancer cell lines, HCC1428 and BT474, with regard to the effect of MIR2052HG on ERα protein levels and cell proliferation (Supplementary Fig. S4). Also, it is well known that androstenedione increases MCF7/AC1 proliferation and that this increase in proliferation can be abrogated with AIs (29). However, overexpression of MIR2052HG significantly increased cell proliferation even in the presence of AI treatment (Fig. 4F). Conversely, down-regulation of MIR2052HG inhibited MCF7/AC1 cell proliferation induced by either E2 or androstenedione (Supplementary Fig. S5). No significant change was observed in the proliferation of ERα-negative cells (AU565 and BT549) after the down-regulation of MIR2052HG (Supplementary Fig. S6).

MIR2052HG regulates ERα expression through transcription and protein degradation pathways

MIR2052HG expression is associated with both ERα mRNA and protein levels (Fig. 4A). To pursue these observations, we began by determining whether MIR2052HG might affect ERα protein stability mediated by protein degradation. Treatment with cycloheximide, resulted in a decrease in the half-life of ERα protein in cells in which MIR2052HG had been knocked down (Fig. 5A and 5B). We also treated MIR2052HG knock down cells with MG132, a proteasome inhibitor, and found that it reversed ERα degradation in these cells (Fig. 5C, upper panel). The same phenomenon was observed using bortezomib, another specific proteasome inhibitor (Fig. 5C, lower panel). These results indicated that MIR2052HG regulates ERα protein stability through a proteasome-mediated degradation pathway. Furthermore, we observed that ERα ubiquitination increased after knocking down

MIR2052HG in 293T cells (Fig. 5D), confirming the involvement of ubiquitin-dependent and proteasome-mediated degradation.

We next examined possible mechanisms by which MIR2052HG was involved in the control of ERα transcription. During these experiments, we observed in both MCF7/AC1 and CAMA1 cells that down-regulation of MIR2052HG resulted in increased phospho-AKT (pAKT) levels at both the Ser473 and Thr308 sites, but total AKT levels did not change (Fig 5E). Because FOXO3 is downstream of AKT and activated AKT phosphorylates FOXO3 resulting in degradation of FOXO3 through a proteasome-dependent process (30), we determined the effects of MIR2052HG knock down on total FOXO3 and phospho-FOXO3 levels (on S318/S321) and observed both of them to be reduced (Fig. 5E), consistent with the known effect of pAKT (30). It is also well known that FOXO3 regulates the expression of ERα (31–33) and that the expression of FOXO3 is directly correlated with the expression of ERα. To further confirm that the regulation of *ESR1* mRNA levels by MIR2052HG is mediated through the regulation of FOXO3, we overexpressed FOXO3 in MCF/AC1 cells, in which MIR2052HG had been knocked down, and observed that FOXO3 overexpression could reverse the down-regulation of *ESR1* mRNA caused by knocking down MIR2052HG (Fig. 5F). In summary, these results indicate that the down-regulation of MIR2052HG can reduce *ESR1* mRNA levels by promoting AKT-mediated down-regulation of FOXO3, which regulates ESR1 transcription. Thus, it appears that tumor expression of MIR2052HG plays a role in the regulation of ERα transcription in addition to ERα protein degradation.

Discussion

Recurrence of breast cancer in women with early stage disease treated with adjuvant endocrine therapy implies endocrine resistance. Multiple potential mechanisms, mainly focusing on ERα function, have been proposed for this resistance but, by and large, these mechanisms have been related to factors present in the cancers (34–36). Much less attention has been paid to host-related factors for endocrine resistance or, more appropriately, lack of efficacy of endocrine therapy, such as CYP2D6 poor metabolizer genotype in the case of tamoxifen (37). In the current study, our goal was to interrogate the germline genome for SNPs related to breast cancer events in women treated with adjuvant AI therapy, relate those SNPs to genes, and perform functional studies to identify potential mechanisms for the observed associations. We controlled for significant imbalances in three stratification factors by performing stratified Cox analyses and for the effects of baseline factors that impacted BCFI with their forced inclusion in the Cox model. The p-values for the top SNPs were approximately 2E-07, which approaches, but does not reach genome-wide significance. However, because of the importance of the phenotype and the fact that we were studying the largest study that had evaluated AIs and had DNA available, we chose to pursue these signals with functional genomic experiments, with strikingly positive findings. We acknowledge that a replication dataset would have been of value but we considered it important to report our compelling data despite the lack of an available dataset.

Our GWAS identified variant SNPs on chromosome 8 that were protective and in or near a gene (*MIR2052HG*) that encodes a lncRNA. There is increasing appreciation of the role of lncRNAs in regulation of the genome (38). For example, lncRNAs can form extensive

networks of ribonucleoprotein complexes with chromatin regulators and modulate them (39, 40). There is also increasing evidence suggesting lncRNAs play important roles in cancers (41, 42). For example, HOTAIR (Hox antisense intergenic RNA) is highly induced in about one-quarter of patients with breast cancer (43). The long intergenic non-coding RNA-ROR has been shown to induce epithelial-to-mesenchymal transition and contribute to breast cancer metastasis (44) and to enhance ERα signaling, conferring resistance to tamoxifen (45). A recent study also indicated that a cluster of lncRNAs, termed Eleanors (ESR1 locus enhancing and activating noncoding RNAs) located within the genomic region containing the ESR1 gene can regulate ERα levels through an enhancer function (46). Of note is the fact that the lncRNA SRA1 (steroid receptor RNA activator 1) acted as a coactivator of ERα, and this action depended on the phosphorylation of ERα at Ser118 (47).

The role of lncRNAs in resistance to endocrine therapy is an area of emerging interest (48, 49). However, our study is, to our knowledge, the first to focus on the impact of a lncRNA on outcomes in a large prospective trial of AI therapy in women with early-stage breast cancer. Our finding of the relationship between SNPs related to the *MIR2052HG* lncRNA and recurrence of breast cancer in women treated with adjuvant AIs led us to perform a series of functional studies to discover potential mechanisms for this association. Using an LCL model system, we showed that, in the presence of E2, these variant SNPs increased both MIR2052HG and ESR1 expression with increased ERα binding to ERE motifs for variant SNP genotypes as shown by ChIP assays (Figs. 2C, 2D). However, when an AI (exemestane or anastrozole) was added, LCLs with the WT SNP genotype displayed a marked up-regulation of MIR2052HG expression and, in parallel, ESR1 expression whereas cells with the variant SNP genotypes showed a clear decrease in both MIR2052HG and ESR1 expression (Fig. 3). That is, the presence of the AI brought about a “reversal” of the SNP-dependent MIR2052HG expression pattern. SNP- and drug-dependent regulation of gene expression has been previously reported by our group in the case of selective estrogen receptor modulators (SERMs), tamoxifen and raloxifene, when given as preventive therapy, which led to the identification of novel mechanisms by which SNPs can regulate gene expression (25). Thus, our previous observations with SERMs (25) provided an impetus to investigate the role of this lncRNA in the efficacy of AI therapy as well as in the regulation of hormone-dependent breast cancer.

MIR2052HG is located on chromosome 8 and not in the *ESR1* genomic region on chromosome 6, but it has a significant effect on ERα regulation as noted above. ERα plays an essential role in cell proliferation and survival in estrogen-dependent breast cancers and in AI-treated patients, *ESR1* amplification, resulting in increased ERα expression, is associated with endocrine resistance (36). We showed that overexpression of MIR2052HG increased ERα expression and accelerated cell proliferation of MCF/AC1 cells in the presence of AI treatments (Figs. 4E, 4F). Conversely, down-regulation of MIR2052HG reduced cell proliferation and colony formation even after E2 or androstenedione treatment (Figs. 4C, 4D and Supplementary Fig. 5). These phenomena are specific for ERα-positive cells, since in ER-negative cell lines, we did not observe an effect of MIR2052HG on cell proliferation (Supplementary Fig. 6). Based on our observations of the effects of MIR2052HG on both ERα protein and mRNA levels, we hypothesized that MIR2052HG might regulate ERα through a proteasome-mediated pathway, which we experimentally

confirmed by treatment with proteasome inhibitors and with ubiquitin assays (Fig. 5). Furthermore, regulation of the transcription of ERα by MIR2052HG was found to be through AKT-dependent FOXO3 regulation; FOXO3 is a known transcription factor for ERα (Figs. 5E, 5F).

Through our GWAS using germline DNA samples from the largest AI clinical trial (5), we have identified a novel lncRNA that potentially plays an important role in the regulation of ERα levels, one of the mechanisms involved in AI resistance. Our GWAS indicated that two variant SNPs (rs4476990 and rs3802201) were common variants with MAF values of 42% and 32%, respectively, and that both were protective, i.e. patients with the variant SNP genotypes had a longer BCFI. The results of our mechanistic studies supported the association in that both variant SNPs down-regulated MIR2052HG expression in the presence of AIs, which was associated with the down-regulation of ERα at both the mRNA level and the protein levels. The presence of markedly increased ESR1 expression in the presence of either anastrozole or exemestane in LCLs with the WT SNP genotype is a potential mechanism for the adverse outcomes in patients carrying the WT SNP genotype. Conversely, down-regulation of ERα in the presence of the variant SNP genotypes after exposure to anastrozole or exemestane might be a factor contributing to their more favorable BCFI.

At the mechanistic level, how MIR2052HG regulates the AKT pathway remains to be further investigated. It could have a direct impact on AKT phosphorylation or more likely, through the regulation of upstream proteins that affect AKT activity. Additionally, the mechanism by which MIR2052HG regulates ERα ubiquitin- and proteasome-mediated degradation also remains unresolved, but previous studies have suggested ER phosphorylation can influence its ubiquitination (50). Therefore, one possibility is that the MIR2052HG could affect ERα levels by regulating various proteins that might affect ERα phosphorylation. Our current findings suggest that MIR2052HG could affect both ERα mRNA and protein levels through different mechanisms.

In summary, we have identified SNP genotypes on chromosome 8 that were associated with breast cancer outcomes in women treated with the AIs anastrozole or exemestane as adjuvant therapy for their early-stage breast cancer, and we related these SNPs to a lncRNA, which, in a SNP-dependent and AI-dependent fashion, regulated ERα expression. The variant SNP genotypes, which are favorable in women treated with anastrozole and exemestane, are potential markers that could identify women for whom either AI would be appropriate therapy, i.e., for those women whose germline carries the variant SNP genotypes. This would require further corroboration with additional clinical testing. However, our GWAS and functional studies have provided initial evidence that germline genetic variability in SNP genotypes related to a gene encoding a lncRNA, *MIR2052HG*, may impact outcomes after treatment with the AIs anastrozole or exemestane.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: These studies were supported in part by NIH grants U19 GM61388 (The Pharmacogenomics Research Network), P50CA116201 (Mayo Clinic Breast Cancer Specialized Program of Research Excellence), U10CA77202, R01CA196648, CCS 015469 from the Canadian Cancer Society, the Breast Cancer Research Foundation, and the RIKEN Center for Integrative Medical Sciences and the Biobank Japan Project funded by the Ministry of Education, Culture, Sports, Science and Technology, Japan. The MA.27 trial was supported, in part, by Pfizer, Inc.

The authors acknowledge the women who participated in the MA.27 clinical trial and provided DNA and consent for its use in genetic studies.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin. 2015; 65:5–29. [PubMed: 25559415]
2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015; 136:E359–86. [PubMed: 25220842]
3. Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet. 2011; 378:771–84. [PubMed: 21802721]
4. Dowsett M, Forbes JF, Bradley R, Ingle J, Aihara T, Bliss J, et al. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. Lancet. 2015; 386:1341–52. [PubMed: 26211827]
5. Goss PE, Ingle JN, Pritchard KI, Ellis MJ, Sledge GW, Budd GT, et al. Exemestane versus anastrozole in postmenopausal women with early breast cancer: NCIC CTG MA.27--a randomized controlled phase III trial. J Clin Oncol. 2013; 31:1398–404. [PubMed: 23358971]
6. Hudis CA, Barlow WE, Costantino JP, Gray RJ, Pritchard KI, Chapman JA, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. J Clin Oncol. 2007; 25:2127–32. [PubMed: 17513820]
7. Ingle JN, Schaid DJ, Goss PE, Liu M, Mushiroda T, Chapman JA, et al. Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. J Clin Oncol. 2010; 28:4674–82. [PubMed: 20876420]
8. Liu M, Goss PE, Ingle JN, Kubo M, Furukawa Y, Batzler A, et al. Aromatase inhibitor-associated bone fractures: a case-cohort GWAS and functional genomics. Mol Endocrinol. 2014; 28:1740–51. [PubMed: 25148458]
9. Sicotte, HNPN.; Johnson, JA.; Lin, Y.; Decker, PA.; Eckel-Passow, JE.; Jenkins, RB.; Goetz, MPR.; Weinshilboum, R.; Boughey, JC.; Wang, L.; Kalari, KR.; McDonnell, SK.; de Andrade, M.; Kocher, JP. Single Sample Imputation from Next Generation Sequencing (NGS) Exome data can improve genotypes in low-coverage regions. 65th Annual Meeting of The American Society of Human Genetics; Baltimore, MD. 2015;
10. PROC PHREG : : SAS Institute Inc. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc; 2008.
11. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2:e190. [PubMed: 17194218]
12. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]
13. Therneau, TM. A Package for Survival Analysis in S. version 2.38. 2015. Retrieved from http://CRAN.R-project.org/package=survival
14. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]

15. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet. 2002; 70:425–34. [PubMed: 11791212]

16. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–8. [PubMed: 21478889]

17. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–303. [PubMed: 20644199]

18. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013; 43:1101–33.

19. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012; 6:80–92. [PubMed: 22728672]

20. Kocher JP, Quest DJ, Duffy P, Meiners MA, Moore RM, Rider D, et al. The Biological Reference Repository (BioR): a rapid and flexible system for genomics annotation. Bioinformatics. 2014; 30:1920–2. [PubMed: 24618464]

21. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. PLoS Med. 2012; 9:e1001216. [PubMed: 22675273]

22. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet. 2008; 4:e1000072. [PubMed: 18464913]

23. Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, et al. Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet. 2004; 36:40–5. [PubMed: 14702039]

24. Ho MF, Bongartz T, Liu M, Kalari KR, Goss PE, Shepherd LE, et al. Estrogen, SNP-Dependent Chemokine Expression and Selective Estrogen Receptor Modulator Regulation. Mol Endocrinol. 2016; 30:382–98. [PubMed: 26866883]

25. Ingle JN, Liu M, Wickerham DL, Schaid DJ, Wang L, Mushiroda T, et al. Selective estrogen receptor modulators and pharmacogenomic variation in ZNF423 regulation of BRCA1 expression: individualized breast cancer prevention. Cancer Discov. 2013; 3:812–25. [PubMed: 23764426]

26. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]

27. Li L, Fridley B, Kalari K, Jenkins G, Batzler A, Safgren S, et al. Gemcitabine and cytosine arabinoside cytotoxicity: association with lymphoblastoid cell expression. Cancer Res. 2008; 68:7050–8. [PubMed: 18757419]

28. Macedo LF, Guo Z, Tilghman SL, Sabnis GJ, Qiu Y, Brodie A. Role of androgens on MCF-7 breast cancer cell growth and on the inhibitory effect of letrozole. Cancer Res. 2006; 66:7775–82. [PubMed: 16885381]

29. Yue W, Zhou D, Chen S, Brodie A. A new nude mouse model for postmenopausal breast cancer using MCF-7 cells transfected with the human aromatase gene. Cancer Res. 1994; 54:5092–5. [PubMed: 7923123]

30. Giamas G, Filipovic A, Jacob J, Messier W, Zhang H, Yang D, et al. Kinome screening for regulators of the estrogen receptor identifies LMTK3 as a new therapeutic target in breast cancer. Nat Med. 2011; 17:715–9. [PubMed: 21602804]

31. Guo S, Sonenshein GE. Forkhead box transcription factor FOXO3a regulates estrogen receptor alpha expression and is repressed by the Her-2/neu/phosphatidylinositol 3-kinase/Akt signaling pathway. Mol Cell Biol. 2004; 24:8681–90. [PubMed: 15367686]

32. Zou Y, Tsai WB, Cheng CJ, Hsu C, Chung YM, Li PC, et al. Forkhead box transcription factor FOXO3a suppresses estrogen-dependent breast cancer cell proliferation and tumorigenesis. Breast Cancer Res. 2008; 10:R21. [PubMed: 18312651]

33. Morelli C, Lanzino M, Garofalo C, Maris P, Brunelli E, Casaburi I, et al. Akt2 inhibition enables the forkhead transcription factor FoxO3a to have a repressive role in estrogen receptor alpha transcriptional activity in breast cancer cells. Mol Cell Biol. 2010; 30:857–70. [PubMed: 19933843]
34. Musgrove EA, Sutherland RL. Biological determinants of endocrine resistance in breast cancer. Nat Rev Cancer. 2009; 9:631–43. [PubMed: 19701242]
35. Osborne CK, Schiff R. Mechanisms of endocrine resistance in breast cancer. Annu Rev Med. 2011; 62:233–47. [PubMed: 20887199]
36. Ma CX, Reinert T, Chmielewska I, Ellis MJ. Mechanisms of aromatase inhibitor resistance. Nat Rev Cancer. 2015; 15:261–75. [PubMed: 25907219]
37. Schroth W, Goetz MP, Hamann U, Fasching PA, Schmidt M, Winter S, et al. Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. Jama. 2009; 302:1429–36. [PubMed: 19809024]
38. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015; 47:199–208. [PubMed: 25599403]
39. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012; 81:145–66. [PubMed: 22663078]
40. Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. BMC Syst Biol. 2014; 8:83. [PubMed: 25033876]
41. Tsai MC, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. Cancer Res. 2011; 71:3–7. [PubMed: 21199792]
42. Cheetham SW, Gruhl F, Mattick JS, Dinger ME. Long noncoding RNAs and the genetics of cancer. Br J Cancer. 2013; 108:2419–25. [PubMed: 23660942]
43. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010; 464:1071–6. [PubMed: 20393566]
44. Hou P, Zhao Y, Li Z, Yao R, Ma M, Gao Y, et al. LincRNA-ROR induces epithelial-to-mesenchymal transition and contributes to breast cancer tumorigenesis and metastasis. Cell Death Dis. 2014; 5:e1287. [PubMed: 24922071]
45. Xue X, Yang YA, Zhang A, Fong KW, Kim J, Song B, et al. LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. Oncogene. 2016; 35:2746–55. [PubMed: 26364613]
46. Tomita S, Abdalla MO, Fujiwara S, Matsumori H, Maehara K, Ohkawa Y, et al. A cluster of noncoding RNAs activates the ESR1 locus during breast cancer adaptation. Nat Commun. 2015; 6:6966. [PubMed: 25923108]
47. Deblois G, Giguere V. Ligand-independent coactivation of ERalpha AF-1 by steroid receptor RNA activator (SRA) via MAPK activation. J Steroid Biochem Mol Biol. 2003; 85:123–31. [PubMed: 12943696]
48. Hayes EL, Lewis-Wambi JS. Mechanisms of endocrine resistance in breast cancer: an overview of the proposed roles of noncoding RNA. Breast Cancer Res. 2015; 17:40. [PubMed: 25849966]
49. Adams BD, Anastasiadou E, Esteller M, He L, Slack FJ. The Inescapable Influence of Noncoding RNAs in Cancer. Cancer Res. 2015; 75:5206–10. [PubMed: 26567137]
50. Reid G, Denger S, Kos M, Gannon F. Human estrogen receptor-alpha: regulation by synthesis, modification and degradation. Cell Mol Life Sci. 2002; 59:821–31. [PubMed: 12088282]

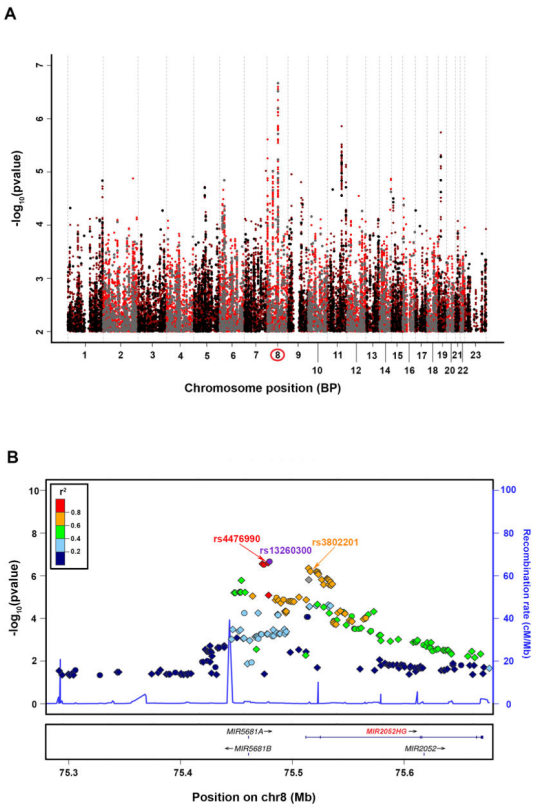

Figure 1.

A, GWAS Manhattan Plot. **B**, Locus zoom of the chromosome 8 region surrounding the *MIR2052HG* gene.

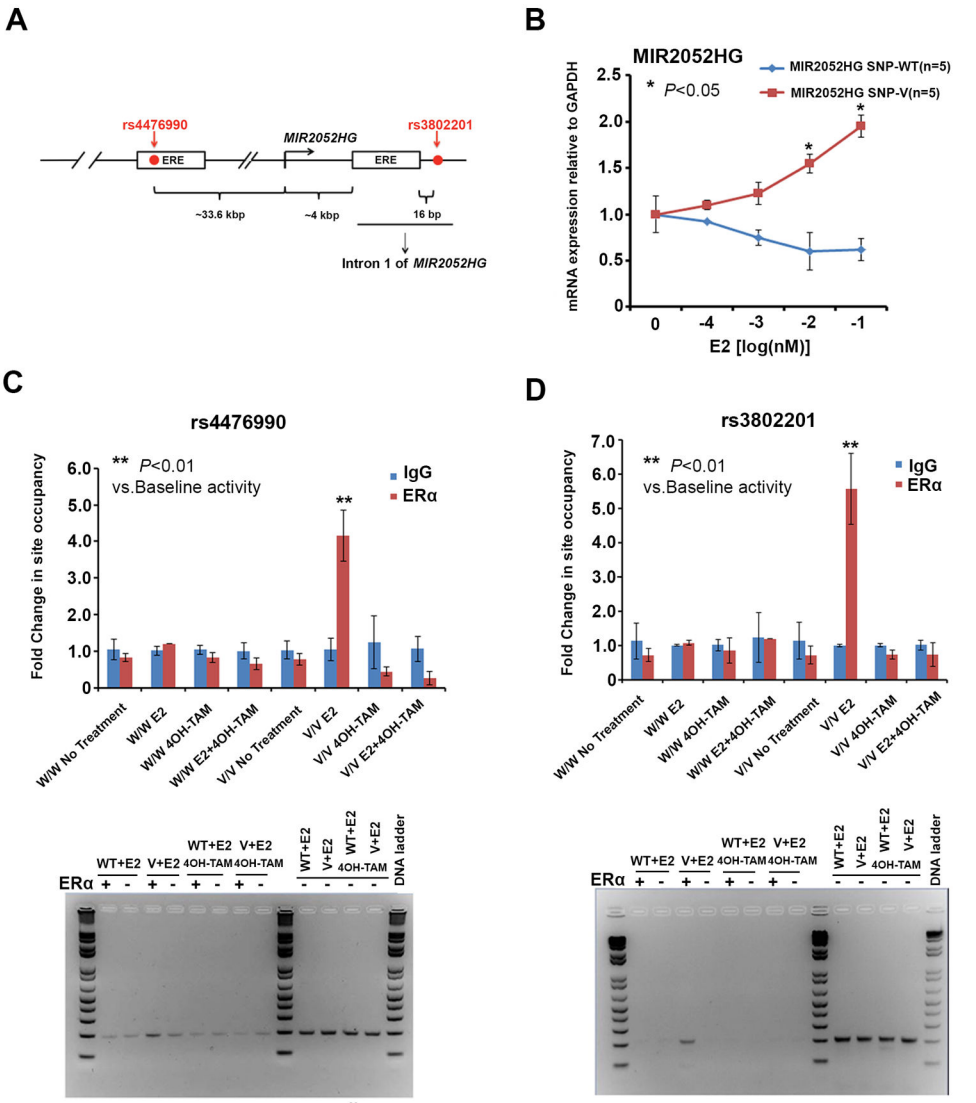

Figure 2.

A. Schematic of estrogen response elements (EREs) around rs4476990 and rs3802201. The EREs are indicated as boxes and the SNPs are indicated as red circles. **B**. MIR2052HG mRNA expression in lymphoblastoid cell lines (LCLs) with wild type SNP (W) and variant SNP (V) genotypes for both rs4476990 and rs3802201 after exposure to increasing concentrations of E2. Error bars represent SEM. *P<0.05. **C.** and **D.**, ChIP assay using six LCLs with known genotypes for rs4476990 and rs3802201 SNPs. Error bars represent SEM of three independent experiments. Representative PCR products are visualized on agarose DNA gels.

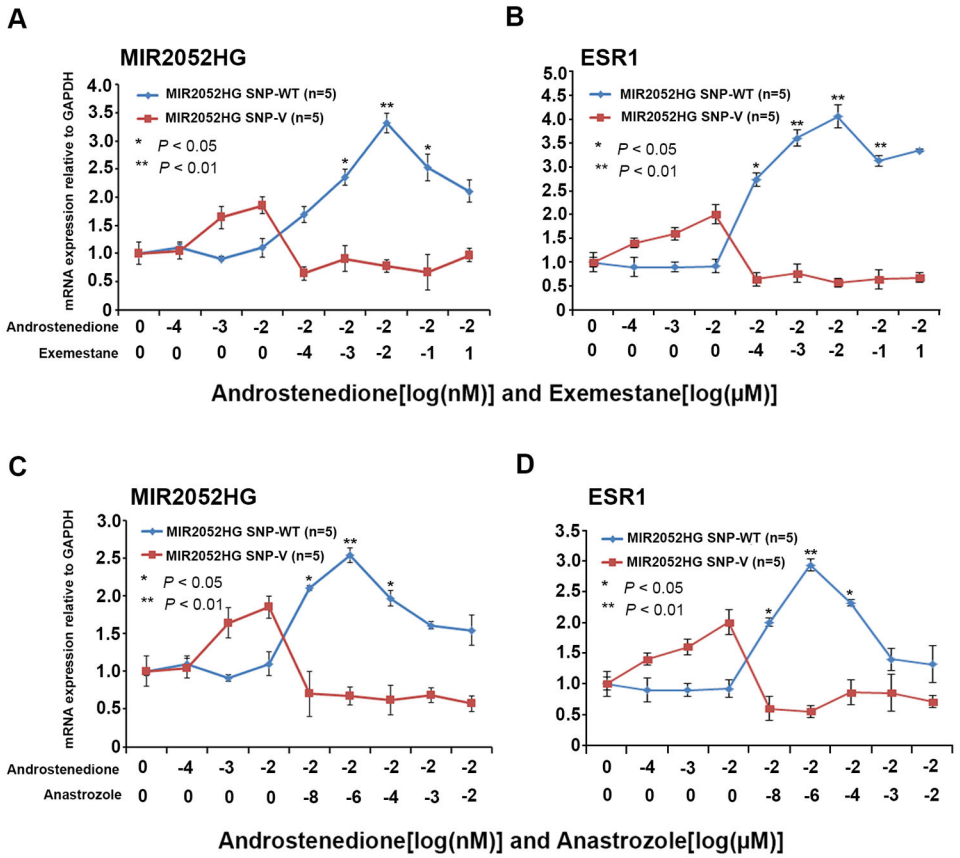

Figure 3.

A and **B**. MIR2052HG mRNA expression in lymphoblastoid cell lines (LCLs) withWT and V genotypes for both rs4476990 and rs3802201 after exposure to androstenedione alone and with increasing concentrations of exemestane or anastrozole. **C** and **D**. mRNA expression for ESR1 in LCLs with the same conditions as A and B. *P<0.05, **P<0.01.

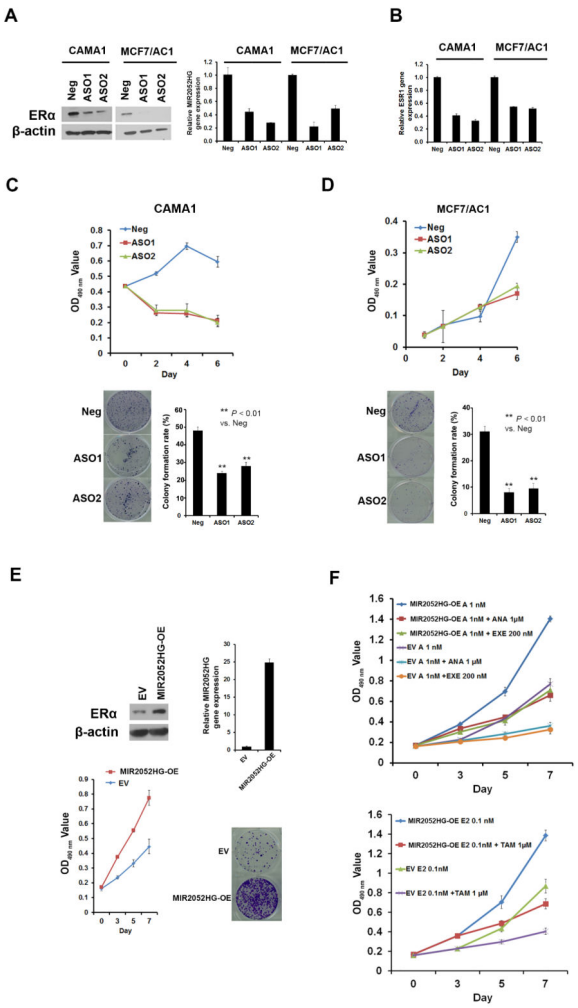

Figure 4.

A, Knock down of MIR2052HG by antisense oligonucleotides (ASO1 and ASO2) down-regulated ERα protein. The histogram shows knock down efficiency in CAMA1 and MCF7/AC1 cells. **B**, Knock down of MIR2052HG decreased ESR1 mRNA expression levels in CAMA1 and MCF7/AC1 cell lines. **C** and **D**. Knock down of MIR2052HG decreased proliferation and colony formation in CAMA1 and MCF7/AC1 cells. The representative colony formation pictures from triplicate experiments are shown. The colony formation rates are quantified as percentages. Error bars represent SEM; ** P< 0.01 compared to baseline (negative control). **E**, Overexpression of MIR2052HG increased ERα protein levels. Overexpression efficiency was determined by qRT-PCR. Overexpression of MIR2052HG increased the proliferation and colony formation and ERα protein levels in MCF7/AC1 cells. **F.** Overexpression of MIR2052HG conferred resistance to AIs (exemestane and anastrozole) and 4-hydroxy-tamoxifen treatments in MCF7/AC1 compared to negative controls. The assay was performed as described in C and D. Error bars represent SEM. The concentrations for androstenedione (A), exemestane (EXE, anastrozole (ANA), estradiol (E2) and 4-hydroxy-tamoxifen (TAM) are indicated.

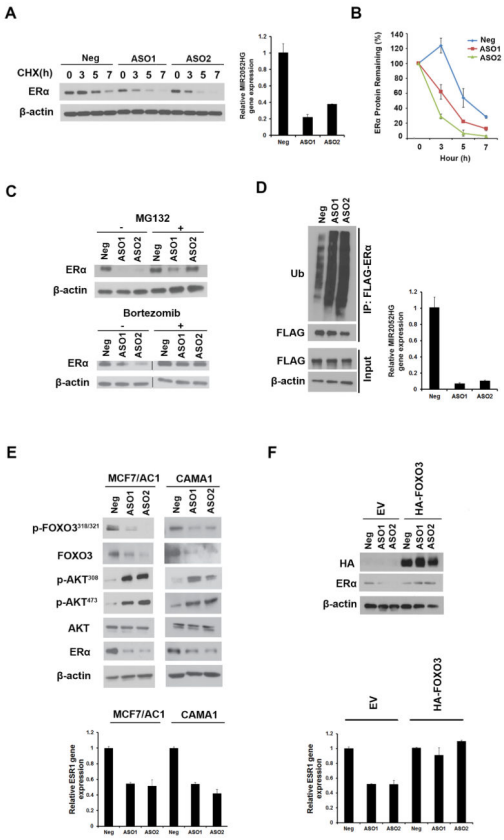

Figure 5.

A, Knock down of MIR2052HG shortened ERα protein half-life. CAMA1 cells were transfected with *MIR2052HG* specific antisense oligonucleotides (ASOs) or a negative control ASO and then treated with cycloheximide (CHX). The representative western blotting results from three independent experiments are shown. The knock down efficiency was determined by qRT-PCR. **B**. Quantitative intensities of ERα are mean values with SEM (error bars) from three independent experiments. **C**. MIR2052HG regulated ERα stability in a proteasome-dependent manner. MIR2052HG was knocked down with ASOs in CAMA1 cells that were treated with MG132 or bortezomib. **D**. Knock down of MIR2052HG promoted the ubiquitination of ERα. 293T cells were transfected with HA-Ub plasmid and FLAG-ERα plasmid, and then transfected with either the MIR2052HG specific ASOs or the negative control ASO followed by MG132. ERα proteins were immunoprecipitated and analyzed by western blotting. Knock down efficiency in 293T cells was determined by qRT-PCR. **E**. MIR2052HG regulated ERα transcription through the AKT-FOXO3 pathway. Knock down of MIR2052HG increased AKT phosphorylation and decreased FOXO3 phosphorylation and FOXO3 total level in MCF7/AC1 and CAMA1 cells. **F**. Overexpression of FOXO3 in MIR2052HG knocked-down MCF7/AC1 cells reversed ERα protein and mRNA levels. Overexpression of FOXO3 was determined by western blotting. HA: expression tag.

Table 1

Patient Characteristics

	Breast Event=No (n=4406)	Breast Event=Yes (n=252)	Total (n=4658)
Genotyped Cohort			
Cohort 1	826 (18.7%)	43 (17.1%)	869 (18.7%)
Cohort 2	834 (18.9%)	47 (18.7%)	881 (18.9%)
Cohort 3	2746 (62.3%)	162 (64.3%)	2908 (62.4%)
Age at Randomization			
Median	64.4	64	64.4
Q1, Q3	58.3, 71.4	58.5, 71.3	58.4, 71.4
Range	36.1 – 95.1	35.9 – 93.6	35.9 – 95.1
Genotypic Race			
African American	134 (3.0%)	6 (2.4%)	140 (3.0%)
Caucasian	4216 (95.7%)	243 (96.4%)	4459 (95.7%)
Asian	56 (1.3%)	3 (1.2%)	59 (1.3%)
Treatment Arm			
Anastrozole	2210 (50.2%)	127 (50.4%)	2337 (50.2%)
Exemestane	2196 (49.8%)	125 (49.6%)	2321 (49.8%)
Adjuvant Chemotherapy			
No	3143 (71.3%)	124 (49.2%)	3267 (70.1%)
Yes	1263 (28.7%)	128 (50.8%)	1391 (29.9%)
ECOG Performance Score			
0	3694 (83.8%)	190 (75.4%)	3884 (83.4%)
1+	712 (16.2%)	62 (24.6%)	774 (16.6%)
T-Stage			
T1/TX	3317 (75.3%)	120 (47.6%)	3437 (73.8%)
T2	997 (22.6%)	115 (45.6%)	1112 (23.9%)
T3	92 (2.1%)	17 (6.8%)	109 (2.3%)
Nodal Status			
Negative (N0)	3290 (74.7%)	121 (48.0%)	3411 (73.2%)
Positive (N1-3)	1037 (23.5%)	129 (51.2%)	1166 (25%)
Unknown (NX)	79 (1.79%)	2 (0.8%)	81 (1.7%)
BMI at baseline			
Median	28.5	28.4	28.5
Q1, Q3	25.0, 32.8	24.9, 32.4	25.0, 32.8
Range	15.3 – 68.7	16.0 – 56.8	15.3 – 68.7
Estrogen receptor, progesterone receptor			
Positive, Positive	3553 (80.6%)	181 (71.8%)	3734 (80.2%)
Positive, Negative	732 (16.6%)	61 (24.2%)	793 (17%)
Positive, Missing	99 (2.2%)	6 (2.4%)	105 (2.2%)
Negative, Positive	22 (0.5%)	4 (1.6%)	26 (0.6%)
Bisphosphonate Use			

	Breast Event=No (n=4406)	Breast Event=Yes (n=252)	Total (n=4658)
No	2949 (66.9%)	196 (77.8%)	3145 (67.5%)
Yes	1457 (33.1%)	56 (22.2%)	1513 (32.5%)
Trastuzumab use			
No	859 (19.5%)	27 (10.7%)	886 (19%)
Unknown	3507 (79.6%)	223 (88.5%)	3730 (80.1%)
Yes	40 (0.9%)	2 (0.8%)	42 (0.9%)
Type of Recurrence			
Any Distant		170 (67.5%)	
Local/Regional +/− Contra (no distant)		42 (16.7%)	
Contralateral Only		40 (15.9%)	

Table 2

SNPs with Lowest P-Values on Chromosome 8

Reference sequence	Base Position	Hazard Ratio* (95% CI)	P-Value	Minor Allele Frequency	Genotyped or Imputed
rs13260300	75476924	0.61 (0.50–0.74)	2.15E-07	0.39	O†
rs4476990§	75478412	0.61 (0.51–0.74)	2.51E-07	0.42	I‡
rs2891356	75473223	0.61 (0.51–0.74)	2.76E-07	0.39	I
rs746460	75474529	0.62 (0.51–0.75)	3.05E-07	0.39	O
rs4735715¶	75514554	0.59 (0.48–0.73)	4.47E-07	0.32	I
rs3802201§,¶	75516199	0.60 (0.48–0.74)	6.24E-07	0.32	I

*For event in patients carrying the variant SNP genotype;

†O: genotyped in at least one cohort;

‡I: imputed in all three cohorts;

§SNP in or near (within 500 bp) an estrogen response element motif;

¶SNP in *MIR2052HG*.